

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2002年 9月17日

出 願 番 号

Application Number:

特願2002-270318

[ ST.10/C ]:

[ JP2002-270318 ]

出 願 人

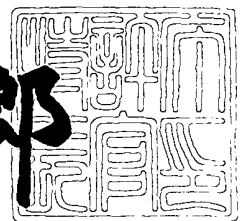
Applicant(s):

株式会社東芝

2003年 3月14日

特 許 庁 長 官  
Commissioner,  
Japan Patent Office

太田信一郎



出証番号 出証特2003-3016613

【書類名】 特許願

【整理番号】 13B0290191

【提出日】 平成14年 9月17日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 11/00

【発明の名称】 指向性設定装置、指向性設定方法及び指向性設定プログラム

【請求項の数】 16

【発明者】

    【住所又は居所】 神奈川県川崎市幸区小向東芝町 1 番地 株式会社東芝  
    研究開発センター内

    【氏名】 天田 皇

【発明者】

    【住所又は居所】 東京都青梅市新町 3 丁目 3 番地の 1 東芝デジタルメデ  
    ィアエンジニアリング株式会社内

    【氏名】 山本 琢己

【特許出願人】

    【識別番号】 000003078

    【住所又は居所】 東京都港区芝浦一丁目 1 番 1 号

    【氏名又は名称】 株式会社東芝

【代理人】

    【識別番号】 100076233

    【弁理士】

    【氏名又は名称】 伊藤 進

【手数料の表示】

    【予納台帳番号】 013387

    【納付金額】 21,000円

【提出物件の目録】

    【物件名】 明細書 1

【物件名】	図面	1
【物件名】	要約書	1
【プルーフの要否】	要	

【書類名】 明細書

【発明の名称】 指向性設定装置、指向性設定方法及び指向性設定プログラム

【特許請求の範囲】

【請求項 1】 複数のマイクロフォンによって構成され入力音響を取込むマイクロフォンアレイと、

前記入力音響に基づく音響信号から特定のキーワードを検出して前記特定のキーワードの音響信号の発生時間に基づく指向性判定期間を求める音声認識手段と

、  
前記指向性判定期間における前記入力音響の音源方向を検出して検出結果を出力する指向性検出手段と、

前記指向性検出手段の検出結果に基づいて所定の機器の指向性を形成する指向性形成手段とを具備したことを特徴とする指向性設定装置。

【請求項 2】 前記指向性形成手段は、前記方向検出手段の検出結果に基づいて前記マイクロフォンアレイの指向性を形成することを特徴とする請求項 1 に記載の指向性設定装置。

【請求項 3】 前記方向検出手段は、前記入力音響の音源方向を検出する方向検出手段と、

前記方向検出手段の検出結果を記憶する記憶手段と、

前記記憶手段から前記指向性判定期間に対応した期間の検出結果を読み出して前記入力音響の音源方向を特定する指向性決定手段とを具備したことを特徴とする請求項 1 に記載の指向性設定装置。

【請求項 4】 前記方向検出手段は、前記入力音響を記憶する記憶手段と、

前記記憶手段から前記指向性判定期間に対応した期間の入力音響を読み出して前記入力音響の音源方向を特定する指向性決定手段とを具備したことを特徴とする請求項 1 に記載の指向性設定装置。

【請求項 5】 前記音声認識手段は、前記複数のマイクロフォンのうちの所定のマイクロフォンからの入力音響から特定のキーワードを検出することを特徴とする請求項 1 に記載の指向性設定装置。

【請求項 6】 前記音声認識手段は、前記指向性形成手段の出力から特定のキーワードを検出することを特徴とする請求項 2 に記載の指向性設定装置。

【請求項 7】 前記指向性検出手段は、前記入力音響の音源方向の検出を複数回繰返すことを特徴とする請求項 1 に記載の指向性設定装置。

【請求項 8】 前記音声認識手段は、前記指向性判定期間として前記特定のキーワードの音響信号の発生時間内の所定期間を設定することを特徴とする請求項 1 に記載の指向性設定装置。

【請求項 9】 前記音声認識手段は、前記指向性判定期間として前記特定のキーワードの音響信号の発生時間内の所定レベル以上の期間を設定することを特徴とする請求項 1 に記載の指向性設定装置。

【請求項 10】 前記指向性形成手段は、複数の機器の指向性を独立して制御可能であることを特徴とする請求項 1 に記載の指向性設定装置。

【請求項 11】 前記音声認識手段は、前記指向性形成手段による指向性形成する処理を解除するためのキーワードも認識して指向性解除信号を出力可能であり、

前記指向性形成手段は、前記音声認識手段からの指向性解除信号が与えられると、前記所定の機器の指向性を解除することを特徴とする請求項 1 に記載の指向性設定装置。

【請求項 12】 前記指向性形成手段は、前記指向性解除信号の入力前に前記指向性検出手段の検出結果が与えられた場合には、前記指向性検出手段の検出結果を記憶し、前記指向性解除信号の入力後に記憶した前記指向性検出手段の検出結果に基づいて所定の機器の指向性を形成することを特徴とする請求項 11 に記載の指向性設定装置。

【請求項 13】 前記指向性検出手段は、前記入力音響の音源方向の検出を複数回繰返すことを特徴とする請求項 2 に記載の指向性設定装置。

【請求項 14】 前記指向性検出手段は、1 回目の前記入力音響の音源方向の検出結果に基づいて前記指向性形成手段が指向性を形成した前記マイクロフォンアレイの出力を前記音声認識手段に与えることで、前記特定のキーワード以外の音声を用いて前記入力音響の音源方向の 2 回目以降の検出を実施することを特徴

とする請求項 7 に記載の指向性設定装置。

【請求項 1 5】 複数のマイクロフォンによって構成されたマイクロフォンアレイによって入力音響を取込む処理と、

前記入力音響に基づく音響信号から特定のキーワードを検出して前記特定のキーワードの音響信号の発生時間に基づく指向性判定期間を求める認識処理と、

前記指向性判定期間における前記入力音響の音源方向を検出して検出結果を得る指向性検出処理と、

前記指向性検出処理の検出結果に基づいて所定の機器の指向性を形成する指向性形成処理とを具備したことを特徴とする指向性設定方法。

【請求項 1 6】 コンピュータに、

複数のマイクロフォンによって構成されたマイクロフォンアレイによって入力音響を取込む処理と、

前記入力音響に基づく音響信号から特定のキーワードを検出して前記特定のキーワードの音響信号の発生時間に基づく指向性判定期間を求める認識処理と、

前記指向性判定期間における前記入力音響の音源方向を検出して検出結果を得る指向性検出処理と、

前記指向性検出処理の検出結果に基づいて所定の機器の指向性を形成する指向性形成処理とを実行させることを特徴とする指向性設定プログラム。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】

本発明は、複数のマイクロホンによるマイクロホンアレイを用いることで機器の指向性を形成するようにした指向性設定装置、指向性設定方法及び指向性設定プログラムに関する。

【 0 0 0 2 】

【従来の技術】

近年、音声認識技術の性能向上に伴い、実環境における音声認識エンジンの実用化が活発になってきている。特に、カーナビゲーションシステムやモバイル機器など入力装置が限定されるような状況において、音声認識への期待は大きい。

## 【 0 0 0 3 】

音声認識処理は、マイクロフォンから取込んだ入力音声を、認識対象語彙と比較することで音声認識結果を得る。実環境下においては、種々の雑音源があることから、マイクロフォンで取込んだ音声信号には、環境雑音が混入する。音声認識処理においては、耐雑音性が認識精度に大きな影響を与える。例えば、車内で音声認識を行う場合には、車のエンジン音や、風切り音、対向車や追い越し車両の音、カーステレオの音等の多くの雑音が存在する。これらの雑音は、発声者の声に混ざって声認識装置に入力され、認識率を大きく低下させる原因となる。

## 【 0 0 0 4 】

このような雑音の問題を解決する方法の一つとして、複数のマイクロフォンを用いて雑音を抑圧するマイクロフォンアレイ技術が知られている。マイクロフォンアレイ技術は、複数のマイクロフォンから入力された音声に対して信号処理を行なうことで、目的とする音声の方向に対して鋭い指向性を形成すると共に、その他の方向の感度を下げて、目的音声の強調を実現している。

## 【 0 0 0 5 】

例えば非特許文献 1 に記載された遅延和型のマイクロフォンアレイ（遅延和アレイ）の場合には、その出力信号  $Se(t)$  は、 $N$  本のマイクロフォンで得られた信号  $Sn(t)$  ( $n=1, \dots, N$ ) を、目的音声の到来方向に合わせた時間差  $\tau$  だけずらして加算することで得られる。つまり、強調された音声信号  $Se(t)$  は、下記（1）式によって与えられる。

## 【 0 0 0 6 】

$$Se(t) = \sum_{n=1}^N Sn(t+n\tau) \quad \dots (1)$$

ただし、複数のマイクロフォンは等間隔で添字  $n$  の順で配置されているものとする。

## 【 0 0 0 7 】

遅延和アレイは到来信号の位相差を利用することで目的音声の方向に指向性を形成している。つまり、遅延和アレイでは、到来信号の各マイクロフォンまでの

伝搬時間差 $\tau$ を考慮して各マイクロフォンの入力信号に対する遅延時間を設定する。各信号に対する遅延処理によって目的音の到来方向からの音声信号（目的信号）同士の位相を同相にした後、相互に加算することで目的信号の強調を行う。一方、目的信号と異なる方向から到来した雑音については遅延処理によって位相が相互にずれることになり、雑音成分は相互に弱めあう。

## 【 0 0 0 8 】

このような遅延和アレイでは、目的音声の到来方向（DOA: Direction of Arrival）に対応する時間差 $\tau$ をどのように推定するかが重要である。 $\tau$ の推定を誤ると、遅延後の目的音声同士の位相がずれてしまい、目的音自体が抑圧されてしまい、性能劣化を招来する。このように、マイクロフォンアレイの指向性を利用して雑音抑圧を行う技術においては、DOAの推定が極めて重要である。DOAの推定については、盛んに研究されており、非特許文献1に記載されているように、線形予測法、最小分散法、MUSIC法等さまざまな方法が提案されている。

## 【 0 0 0 9 】

例えば、特許文献1においては、マイクロフォンアレイを用いて、話者方向を逐次検知してその方向にマイクロフォンアレイの指向性を修正することで、話者の方向を追尾し、目的信号に対する歪みを抑圧する方法も開示されている。

## 【 0 0 1 0 】

しかし、例えば複数の人物が発声する場合においては、話者の方向が必ずしも目的音の到来方向であるとは限らない。例えば、複数の人物のうちの特定の人物のみが目的音を発声し、他の人物の発声は雑音であることもある。この場合には、複数の人物の内の特定の人物の方向のみが目的音の到来方向であり、特許文献1ではこの場合に対応することはできない。

## 【 0 0 1 1 】

そこで、DOAの推定を確実なものとするために、特許文献2においては、予め音源領域を設定しキーワードに対応付けて登録する手法が開示されている。即ち、特許文献2においては、予めマイクロフォンアレイに対する各話者の位置（音源領域）をキーワードと共に登録しておく。入力音声からキーワードを認識すると、話者の位置とキーワードとが登録されたテーブルを参照することで、認



識したキーワードに対応する音源領域を特定し、以後、この音源領域に対して鋭い指向性に設定する。これにより、確実なDOAの検出を可能にして、音声認識精度を向上させている。

【0012】

【特許文献1】

特開平9 - 9 7 9 4 号公報

【0013】

【特許文献2】

特開2002 - 3 4 0 9 2 号公報

【0014】

【非特許文献1】

「音響システムとデジタル処理」,第7章,電子情報通信学会,1995)

【0015】

【発明が解決しようとする課題】

特許文献2の方法は、確実にDOAを設定する方法としては有効であるが、設定可能なDOA、即ち、話者の位置が予め固定されてしまう。また、固定の話者の位置をキーワードと共に登録して、記憶させておく必要もあるという問題があった。

【0016】

本発明は、指向性の方向を記憶させることなく、また、指向性の方向を自由に設定可能にすると共に、確実に目的音の到来方向に指向性を設定することができる指向性設定装置、指向性設定方法及び指向性設定プログラムを提供することを目的とする。

【0017】

【課題を解決するための手段】

本発明の請求項1に係る指向性設定装置は、複数のマイクロフォンによって構成され入力音響を取込むマイクロフォンアレイと、前記入力音響に基づく音響信号から特定のキーワードを検出して前記特定のキーワードの音響信号の発生時間に基づく指向性判定期間を求める音声認識手段と、前記指向性判定期間における

前記入力音響の音源方向を検出して検出結果を出力する指向性検出手段と、前記指向性検出手段の検出結果に基づいて所定の機器の指向性を形成する指向性形成手段とを具備したものである。

【 0 0 1 8 】

本発明の請求項 1 において、入力音響はマイクロフォンアレイを介して取込まれる。音声認識手段は入力音響に基づく音響信号から特定のキーワードを検出して、このキーワードの音響信号の発生時間に基づいて指向性判定期間を求める。指向性検出手段は、指向性判定期間における入力音響の音源方向を検出する。この検出結果は指向性形成手段に与えられ、指向性形成手段は、検出された入力音響の音源方向に基づいて、所定の機器の指向性を形成する。

【 0 0 1 9 】

なお、装置に係る本発明は方法に係る発明としても成立する。

【 0 0 2 0 】

また、装置に係る本発明は、コンピュータに当該発明に相当する処理を実行させるためのプログラムとしても成立する。

【 0 0 2 1 】

【発明の実施の形態】

以下、図面を参照して本発明の実施の形態について詳細に説明する。図 1 は本発明の第 1 の実施の形態に係る指向性設定装置を示すブロック図である。

【 0 0 2 2 】

本実施の形態はマイクロフォンアレイを用いた雑音抑圧装置に適用したものであり、発話開始を表す特定のキーワードを設定し、このキーワードを音声認識することによって、キーワード自体の D O A を推定して、その方向に、ある期間だけ指向性を形成するものである。これにより、発話可能な複数の人物のうち、キーワードを発した特定の人物の音声のみを強調する共に他の到来方向の雑音を抑圧する音響処理が可能である。即ち、発話位置やユーザーを限定することなく、不特定多数の人物が任意の位置から発声した場合でも、確実に D O A を検出することができる。

【 0 0 2 3 】

図 1 において、複数のマイクロフォン 1 0 1 -1 ~ 1 0 1 -N は、相互に所定の間隔だけ離間した位置に配置されて、音響信号を取込むマイクロフォンアレイを構成する。複数のマイクロフォン 1 0 1 -1 ~ 1 0 1 -N に取込まれた音響信号は、指向性制御部 2 0 1 及び指向性形成部 2 0 3 に供給される。また、複数のマイクロフォン 1 0 1 -1 ~ 1 0 1 -N のうちの任意の 1 つのマイクロフォン（図 1 の例ではマイクロフォン 1 0 1 -1）からの音響信号は認識部 3 0 0 に供給されている。

#### 【 0 0 2 4 】

認識部 3 0 0 は、入力音響信号に対する音声認識処理が可能であり、予め設定された所定のキーワードの音響がマイクロフォン 1 0 1 -1 を介して入力されたことを検出することができる。認識部 3 0 は、キーワードの音声入力を検出すると、検出結果を指向性制御部 2 0 1 又は指向性形成部 3 0 0 に出力するようになっている。

#### 【 0 0 2 5 】

例えば、認識部 3 0 0 は、入力音響を一定時間間隔（フレーム）毎に音響分析し、分析結果としてケプストラムやパワースペクトル等の音声特徴量の時系列及び音声の開始時刻、終了時刻を得る。認識部 3 0 0 は、所定のキーワードが登録されている図示しないキーワード格納部を有している。認識部 3 0 0 は、分析結果の特徴量とキーワード格納部に登録されている単語の特徴量パターンとの間でパターンマッチングを行って、両者の一致不一致を判定する。

#### 【 0 0 2 6 】

認識部 3 0 0 は、キーワードに相当するフレーム、即ち、発話期間を示す信号を指向性判定期間として指向性制御部 2 0 1 に出力することができるようになっている。なお、認識部 3 0 0 は、キーワードの発話期間のうち入力音響レベルが所定レベル以上の期間のみを指向性判定期間として指定することもできる。更に、認識部 3 0 0 は、キーワードの発話期間よりも所定期間だけ長い期間又は短い期間を指向性判定期間として指定するようにしてもよい。

#### 【 0 0 2 7 】

なお、所定のキーワードとしては、発話開始を表すキーワードや発話終了を表すキーワード等が登録されている。認識部 3 0 0 は、発話開始を表すキーワード

を検出すると指向性判定期間を指向性制御部 2 0 1 に出力し、発話終了を表すキーワードを検出すると検出結果（指向性解除信号）を指向性形成部 2 0 3 に出力する。

## 【 0 0 2 8 】

指向性制御部 2 0 1 は、複数のマイクロフォン 1 0 1 -1 ~ 1 0 1 -N に入力された音響信号から入力音響の到来方向を逐次検出して、入力音響の到来方向を示す方向データを出力する。例えば、指向性制御部 2 0 1 は、フレーム毎に、入力音響の到来方向を検出し、検出結果を逐次指向性記憶部 2 0 2 に出力する。

## 【 0 0 2 9 】

指向性記憶部 2 0 2 は、入力音響の到来方向を示す方向データを逐次記憶するようになっている。指向性記憶部 2 0 2 は、例えばキーワードの発話時間よりも長い時間分、少なくとも指向性判定期間よりも長い時間分の方向データを記憶する記憶容量を有している。

## 【 0 0 3 0 】

指向性制御部 2 0 1 は、指向性記憶部 2 0 2 に方向データを出力して記憶させると共に、指向性記憶部 2 0 2 に蓄積されている方向データを読み出すことができる。指向性制御部 2 0 1 は、認識部 3 0 0 から指向性判定期間を示す信号が与えられると、指向性判定期間に相当する期間の 1 つ以上の方向データを指向性記憶部 2 0 2 から読み出し、読み出した方向データを用いた例えば平均値処理又は最頻値処理等によって、指向性判定期間における入力音響の到来方向を求める。即ち、指向性制御部 2 0 1 は、キーワードの発話者に対するマイクロフォンアレイからの方向を求める。

## 【 0 0 3 1 】

指向性制御部 2 0 1 は、指向性判定期間における入力音響の到来方向の算出結果を到来方向設定信号として指向性形成部 2 0 3 に出力するようになっている。指向性形成部 2 0 3 は、複数のマイクロフォン 1 0 1 -1 ~ 1 0 1 -N からの音響信号が入力され、指向性制御部 2 0 1 からの到来方向設定信号に基づいて、入力された音響信号の指向特性を設定するようになっている。

## 【 0 0 3 2 】

例えば、指向性形成部 2 0 3 は、各マイクロフォン 1 0 1 -1 ~ 1 0 1 -N の設置位置及び到来方向設定信号によって指定された到来方向に基づく遅延量で各マイクロフォン 1 0 1 -1 ~ 1 0 1 -N からの音響信号を遅延させ、各遅延信号同士を加算することで、所望の方向に狭指向性を有する音響信号を得る。即ち、目的音の到来方向の感度は著しく高く、その他の方向の感度は著しく低い処理音を出力する。

#### 【 0 0 3 3 】

また、指向性形成部 2 0 3 は、認識部 3 0 0 から発話終了を示すキーワードを検出したことを示す信号（指向性解除信号）が与えられると、指向性形成処理を停止して、指向性を広指向性（無指向性）に戻すようになっている。なお、指向性形成部 2 0 3 は、発話終了を示すキーワードの検出結果によって指向性形成処理を終了させたが、発話の開始から予め設定された所定時間後に自動的に指向性形成処理を終了するようにしてもよい。

#### 【 0 0 3 4 】

なお、指向性記憶部 2 0 2 は、逐次入力される方向データを記憶するものとして説明したが、指向性制御部 2 0 1 において指向性判定期間における入力音響の到来方向の算出を可能にするデータを記憶させればよく、例えば、到来した音響信号そのものの波形データを記憶するようにしてもよい。この場合には、指向性制御部 2 0 1 は、指向性判定期間が指示された直後に、指向性記憶部 2 0 2 からの波形データを読み出して、指向性判定期間における入力音響の到来方向を求めればよい。

#### 【 0 0 3 5 】

次に、このように構成された実施の形態の動作について図 2 のフローチャートを参照して説明する。

#### 【 0 0 3 6 】

いま、例えば、異なる任意の位置の複数の発話者が 1 台の音声認識装置を利用したアプリケーションを使用するものとする。この場合には、複数の発話者からの音声を図 1 の指向性設定装置のマイクロフォン 1 0 1 -1 ~ 1 0 1 -N から取込み、指向性形成部 2 0 3 からの処理音を音声認識装置に供給する。

## 【0037】

図2のステップS1は発話者からの音声入力の待機状態を示している。マイクロフォン101-1～101-Nは音響入力を取込み、指向性制御部201に供給する。また、マイクロフォン101-1を介して取込まれた音響入力は認識部300に供給される。なお、マイクロフォン101-1～101-Nには目的音の他に雑音も混入した音響信号が入力される。音声認識装置に対して複数の発話者が同時に発話する場合等においては、所定の1人以外の発話者からの音声も雑音となる。

## 【0038】

指向性制御部201は、入力音響に対して、逐次その到来方向を検出する。指向性制御部201が検出した入力音響の到来方向は方向データとして逐次指向性記憶部202に供給されて記憶される（ステップS2）。なお、この場合には、雑音を含む目的音についての到来方向が検出される。

## 【0039】

一方、認識部300は、マイクロフォン101-1から入力される音響信号から発話開始を示すキーワードを認識する。例えば、発話開始を示すキーワードとして「カイン」という語句が設定されるものとする。ここで、全ての発話者のうちの所定の発話者が「カイン」という語句を発声するものとする。この音声はマイクロフォン101-1を介して取込まれて認識部300に供給される。認識部300は、逐次入力される音響に対する所定の音声認識処理によって、「カイン」の語句が発声されたことを検出する。そうすると、指向性制御部201は、この「カイン」の語句の発話期間に基づく指向性判定期間を設定して指向性制御部201に出力する。

## 【0040】

指向性制御部201は、指向性判定期間を示す信号が入力されると、指向性判定期間に相当する期間の方向データを指向性記憶部202から読み出す。例えば、指向性制御部201は、複数の方向データの最頻値によって、指向性判定期間における入力音響（目的音）の到来方向、即ち、所定の発話者（以下、特定発話者という）の方向を決定する（ステップS4）。指向性制御部201はキーワードの到来方向を示す到来方向設定信号を指向性形成部203に出力する。

## 【 0 0 4 1 】

指向性形成部 2 0 3 は、到来方向設定信号が入力されると、マイクロフォン 1 0 1-1～1 0 1-Nを介して入力された音響信号に対する指向特性を、到来方向設定信号に基づいて設定する。これにより、以後、指向性形成部 2 0 3 は、キーワードを発声した特定発話者方向（到来方向）に狭指向性で高い感度を有する指向特性を入力音響に付与した処理音を出力する（ステップ S5 ）。

## 【 0 0 4 2 】

従って、例えば、複数の発話者のうちの特定発話者が「カイン」を発声して、この「カイン」が認識部 3 0 0 においてキーワードとして認識された場合には、以後、特定発話者以外の方向の他の発話者が発話をしている場合でも、或いは、特定発話者以外の方向から雑音が発生している場合であっても、特定発話者の発声のみが強調されて処理音として音声認識装置に供給されることになる。これにより、以後、特定発話者の音声認識精度を著しく向上させることができる。

## 【 0 0 4 3 】

ここで、発話の終了を示すキーワードとして、例えば「シュウリョウ」が設定されているものとする。マイクロフォン 1 0 1-1を介して「シュウリョウ」の音響が認識部 3 0 0 に入力されると、認識部 3 0 0 は、「シュウリョウ」をキーワードとして認識する。そうすると、認識部 3 0 0 は、指向性形成部 2 0 3 に指向性形成処理の終了を指示するための指向性解除信号を出力する。これにより、指向性形成部 2 0 3 は入力音響に対する指向特性の付与を終了する。

## 【 0 0 4 4 】

このように、本実施の形態においては、発話者の数や位置とは無関係の 1 つのキーワードによって発話開始を指示することで、キーワードの発話者の方向を発声されたキーワードによって検出している。これにより、以後、キーワードの特定発話者方向に狭指向性を有する指向特性を設定した処理音が発生している。従って、指向性を形成可能な発話者の位置が固定されてしまうことはなく、また、予め発話者の位置を登録しておく必要もなく、更に、発話者の位置や数に制限を受けることもなく、一旦発話開始のキーワードを発声してこのキーワードが認識されると、以後、確実に所定の期間、又は発話終了のキーワードが発声されるま

で、特定発話者方向からの音声を強調することができ、特定発話者が発生する音声の音声認識精度を著しく向上させることができる。

【 0 0 4 5 】

図 3 は本発明の第 2 の実施の形態に係る指向性設定装置を示すブロック図である。図 3 において図 1 と同一の構成要素には同一符号を付して説明を省略する。

【 0 0 4 6 】

本実施の形態はキーワードの到来方向によってカメラの視野方向を制御するようにしたものである。本実施の形態は指向性形成部 2 0 3 に代えてカメラ駆動制御部 2 1 1 を設けた点が第 1 の実施の形態と異なる。

【 0 0 4 7 】

本実施の形態においては、図示しない複数の発話者方向を撮影可能なカメラが図示しない支持台上に配置されている。そして、支持台は図示しない駆動部によって水平方向の向き及び垂直方向の角度を自由に変更することができるようになっている。カメラ駆動制御部 2 1 1 は、カメラ駆動信号を駆動部に出力することにより、カメラ支持台の水平及び垂直方向の向きを制御することができるようになっている。

【 0 0 4 8 】

本実施の形態においては、カメラ駆動制御部 2 1 1 は、指向性制御部 2 0 1 からの到来方向設定信号に基づいて、カメラ支持台の向きを制御するカメラ駆動信号を出力するようになっている。これにより、カメラは目的音の到来方向を撮影することができるようになっている。即ち、カメラ駆動制御部 2 1 1 は、特定発話者のキーワード発声によって、カメラの視野方向を特定発話者に向ける指向性形成処理を実行する。

【 0 0 4 9 】

また、指向性形成部 2 0 3 は、認識部 3 0 0 から発話終了（指向性解除）を示すキーワードを検出したことを示す信号が与えられると、指向性形成処理を停止して、カメラの視野方向を所定の初期方向に戻すようになっている。なお、指向性形成部 2 0 3 は、発話終了を示すキーワードの検出結果によってカメラの視野方向を特定発話者方向に向ける指向性形成処理を終了させたが、発話の開始から



予め設定された所定時間後に自動的に指向性形成処理を終了させてカメラの視野方向を所定の初期方向に戻すようにしてもよい。

【 0 0 5 0 】

このように構成された実施の形態においては、全発話者中の所定の発話者（特定発話者）が、発話（指向性設定）の開始を示すキーワードを発声し、認識部 3 0 0 によってキーワードの発声が検出されると、以後、所定期間又は発話終了（指向性解除）のキーワードが認識されるまで、カメラの視野方向が特定発話者に自動的に向いて、特定発話者を撮影することが可能となる。

【 0 0 5 1 】

なお、本実施の形態においては、カメラ駆動制御部 2 1 1 は、カメラの視野方向を特定発話者に向けるようにしたが、特定発話者の方向と所定の方向関係を有する方向に視野方向を向けるように設定することも可能である。

【 0 0 5 2 】

また、第 1 及び第 2 の実施の形態を組み合わせることによって、到来方向設定信号に基づいて、マイクロフォン 1 0 1 -1 ~ 1 0 1 -N からの入力音響に対して特定発話者方向に狭指向性を有する指向性を形成すると同時に、カメラに特定発話者方向を撮影させるようにすることができることは明らかである。

【 0 0 5 3 】

図 4 は本発明の第 3 の実施の形態に係る指向性設定装置を示すブロック図である。図 4 において図 1 と同一の構成要素には同一符号を付して説明を省略する。

【 0 0 5 4 】

本実施の形態は認識部 3 0 0 に指向性形成部 2 0 3 からの処理音を供給するようにした点が第 1 の実施の形態と異なる。

【 0 0 5 5 】

本実施の形態においては、初期状態において、例えば指向性形成部 2 0 3 は、広指向性又は無指向性に設定する。これにより、マイクロフォン 1 0 1 -1 ~ 1 0 1 -N からの入力音響は、単に合成されて認識部 3 0 0 に供給される。

【 0 0 5 6 】

このように構成された実施の形態においては、マイクロフォン 1 0 1 -1 ~ 1 0

1-Nによるマイクロフォンアレイからの入力音響に対してキーワードの検出処理を行っており、キーワードの検出結果の精度を向上させることができる。

【0057】

また、初期状態において、指向性形成部203の指向性を所定方向に指向させることも可能である。この場合には、キーワードの検出を可能とする発話者、即ち、特定発話者を限定することも可能である。

【0058】

また、第1の実施の形態においては、認識部300にはマイクロフォン101-1からの音響のみを供給した。このため、発話の終了を示すキーワードを特定発話者方向以外に位置する発話者が発声した場合でも、指向性形成部203による指向性形成処理が終了する。これに対し、本実施の形態においては、認識部300に、指向性形成部203からの処理音を供給していることから、一旦特定発話者が決定すると、以後、特定発話者又は特定発話者方向に位置する発話者が「シュウリョウ」と発声した場合にのみ、このキーワードが認識可能となって、指向性形成処理を終了させることができる。これにより、音声認識装置を用いたアプリケーションの誤動作の可能性を低減することができる。

【0059】

図5は本発明の第4の実施の形態に係る指向性設定装置を示すブロック図である。図5において図1と同一の構成要素には同一符号を付して説明を省略する。

【0060】

本実施の形態は目的音の到来方向の検出を、複数回実行することにより検出精度を向上させるようにしたものである。指向性制御部205は、図1における指向性制御部201と同様に、複数のマイクロフォン101-1～101-Nに入力された音響信号から入力音響の到来方向を逐次検出して、入力音響の到来方向を示す方向データを指向性記憶部202に記憶させることができる。

【0061】

更に、本実施の形態においては、指向性制御部205は、認識部300からの指向性判定期間とは異なる期間の方向データを指向性記憶部202から読み出して、到来方向設定信号を修正することができるようになっている。

## 【 0 0 6 2 】

更に、本実施の形態においては、指向性制御部 2 0 5 は、マイクロフォン 1 0 1-1 ~ 1 0 1-N からの入力音響をそのまま指向性記憶部 2 0 2 に与えて記憶させることもできるようになっている。この場合には、指向性制御部 2 0 5 は、2 回目の到来方向算出時には、指向性記憶部 2 0 2 に記憶されている波形データを読み出して、到来方向の検出演算を再実行するようになっている。2 回目以降の到来方向の算出演算時には、指向性制御部 2 0 5 は、1 回目の到来方向算出演算よりも高い精度での演算を実行するようになっている。

## 【 0 0 6 3 】

なお、指向性記憶部 2 0 2 は保持しているデータを指向性形成部 2 0 3 に供給することができるようになっている。指向性形成部 2 0 3 は、2 回目以降の到来方向の算出演算時には、指向性記憶部 2 0 2 からのデータに対して指向性を形成した処理音を認識部 3 0 0 に供給することができるようになっている。これにより、到来方向の算出演算時に基準となる音声認識結果の情報を 2 回目以降の到来方向算出演算時にも得られるようになっている。

## 【 0 0 6 4 】

1 回目の到来方向の算出演算は、目的音の到来方向として検出すべき方向範囲を大きく設定する必要がある。従って、指向性制御部 2 0 5 による到来方向の算出演算は比較的粗い精度で実施される。これに対し、2 回目以降の到来方向の算出演算は、1 回目の到来方向算出演算によってある程度検出すべき方向範囲が限定されていることから、算出演算を比較的高精度に実施することができる。このように、指向性制御部 2 0 5 は、複数回到来方向の算出演算を実施することで、一層高精度に到来方向を求めることができる。

## 【 0 0 6 5 】

次に、このように構成された実施の形態の動作について説明する。

## 【 0 0 6 6 】

認識部 3 0 0 は、マイクロフォン 1 0 1-1 からの入力音響を用いて、キーワードの検出を行う。認識部 3 0 0 はキーワードを検出すると、発話期間に基づく指向性判定期間を設定して指向性制御部 2 0 5 に指示する。指向性制御部 2 0 5 は

、認識部 3 0 0 からの指向性判定期間の指示に従って、指向性記憶部 2 0 2 に記憶されている方向データ又は波形データから目的音の到来方向を算出する。指向性制御部 2 0 5 は求めた到来方向を示す到来方向設定信号を指向性形成部 2 0 3 に出力する。こうして、指向性形成部 2 0 3 からはキーワードの特定発話者に指向性を有する処理音が出力される。

## 【 0 0 6 7 】

更に、本実施の形態においては、指向性制御部 2 0 5 は、指向性記憶部 2 0 2 に記憶されている方向データ又は波形データを用いて、複数回の到来方向算出演算を実行することができる。例えば、キーワードの発話期間の開始時間近傍及び終了時間近傍においてノイズレベルが高い場合でも、指向性を制御する動作と、その指向性により得られた信号を認識する動作を繰返すことで、雑音を除去し音声のみを取り出せるため、キーワードに対するより正確な発声区間が得られ、到来方向算出演算がキーワードの前後に発声した雑音の影響を受けにくくなり、到来方向算出演算の精度を向上させることができる。

## 【 0 0 6 8 】

また、指向性制御部 2 0 5 は、指向性記憶部 2 0 2 に記憶されている波形データ読み出して、1 回目の到来方向算出時に求めた到来方向近傍の比較的狭い方向範囲について、到来方向の算出を行う。これにより、到来方向算出演算の精度を向上させることができる。

## 【 0 0 6 9 】

指向性制御部 2 0 5 は複数回の到来方向算出演算によって求めた到来方向を示す到来方向設定信号を指向性形成部 2 0 3 に出力する。これにより、指向性形成部 2 0 3 は目的音の到来方向として一層高い精度の狭指向性を設定することができる。

## 【 0 0 7 0 】

このように、本実施の形態においては、1 回の到来方向算出演算の演算結果を利用して 2 回目以降の到来方向算出演算を実行しており、到来方向を一層高精度に検出することが可能である。

## 【 0 0 7 1 】

また、本実施の形態においては、2回目以降の到来方向算出演算において、指向性記憶部202内のデータ、即ち、キーワードを用いた演算を行っている。これに対し、発話者が新たに発話した音響信号に基づいて、2回目以降の到来方向算出演算を実施するようにしてもよい。この場合には、指向性記憶部202内にデータを記憶させておく必要はないが、認識部300において、キーワード以外の音声認識が必要である。

【0072】

図6は本発明の第5の実施の形態に係る指向性設定装置を示すブロック図である。図6において図1と同一の構成要素には同一符号を付して説明を省略する。

【0073】

第1の実施の形態においてはキーワードを発話する1人の特定発話者のみに指向性を形成するようにした。これに対し、本実施の形態においては、キーワードを発話する複数の特定発話者に対して夫々指向性を形成可能にしたものである。

【0074】

本実施の形態においては、複数の指向性形成部203-1～203-Nを設けた点が第1の実施の形態と異なる。指向性形成部203-1～203-Nは、指向性形成部203と同様の構成であり、指向生成制御部201から夫々到来方向設定信号が与えられて、入力された到来方向設定信号に基づいて入力音響の指向性を相互に独立して設定するようになっている。

【0075】

次に、このように構成された実施の形態の動作について説明する。

【0076】

いま、複数の発話者のうち第1の発話者が発話の開始を示すキーワードを発声するものとする。認識部300はこのキーワードを検出して指向性判定期間を指向性制御部201に指示する。これにより、指向性制御部201は、指向性判定期間に相当する期間分の方向データを読み出して、キーワードを発話した第1の発話者（以下、第1の特定発話者）による目的音の到来方向を検出する。この到来方向を示す到来方向設定信号は例えば指向性形成部203-1に供給される。これにより、指向性形成部203-1は、第1の特定発話者方向に狭指向性を有する

指向性を形成する。こうして、指向性形成部 2 0 3-1からの処理音は、第 1 の特定発話者が発声する音声を強調したものとなる。

#### 【 0 0 7 7 】

この状態で、更に、第 2 の発話者が発話の開始を示すキーワードを発声するものとする。認識部 3 0 0 はこのキーワードを検出して指向性判定期間を指向性制御部 2 0 1 に指示する。この場合にも同様に、指向性制御部 2 0 1 は、指向性判定期間に相当する期間分の方向データを読み出してキーワードを発話した第 2 の発話者（以下、第 2 の特定発話者）による目的音の到来方向を検出する。この場合には、指向性制御部 2 0 1 は、到来方向を示す到来方向設定信号を例えば指向性形成部 2 0 3-2 に供給される。これにより、指向性形成部 2 0 3-2 は、第 2 の特定発話者方向に狭指向性を有する指向性を形成する。こうして、指向性形成部 2 0 3-2 からの処理音は、第 2 の特定発話者が発声する音声を強調したものとなる。

#### 【 0 0 7 8 】

このように、本実施の形態においては、複数の指向性形成部を設けることによって、複数の特定発話者方向に同時に指向性を形成することが可能となり、複数の特定発話者が発声した音声を高精度に音声認識処理することを可能にしている。

#### 【 0 0 7 9 】

ところで、第 5 の実施の形態においては、指向性形成部を複数設ける必要があり、装置規模が増大する。そこで、1 個の指向性形成部によって逐次指向性を切替えるようにしてもよい。例えば、指向性制御部 2 0 1 において、キーワードの検出によって到来方向を求めた場合には、前の指向性形成処理が終了した後に、求めた到来方向に基づく到来方向設定信号を指向性形成部に与えるのである。これにより、1 人の特定発話者方向の指向性形成処理が終了した後に、次の特定発話者方向に対する指向性形成処理が実行されることになり、計算量を増大させることなく、各特定発話者の音声認識処理精度を向上させることができる。

#### 【 0 0 8 0 】

図 7 は本発明の第 6 の実施の形態を説明するための説明図である。本実施の形

態は指向性設定装置を車載機器の制御に適用したものである。

【0081】

図7において、車内400には、運転席411及び助手席412が配設されており、運転席411及び助手席412には夫々運転手401及び同乗者402が座っている。車内400の前方の例えば図示しないダッシュボード内には、図示しない車載機器、例えばエアコンの制御部が設置されている。車載機器は音声認識部403の認識結果に応じて動作が制御されるようになっている。

【0082】

そして、音声認識部403には第1、第3乃至第5の実施の形態における指向性設定装置、例えば、第3の実施の形態における指向性設計装置404からの処理音が供給されるようになっている。

【0083】

次に、このように構成された実施の形態の動作について図8を参照して説明する。図8は第6の実施の形態の処理を説明するための説明図である。図8は運転手401と同乗者402の一連の会話と、それに対応する指向性設定装置404と音声認識部403（図8中では指向性設定装置404内の認識部を含む）の動作を時系列で示したものである。なお、図8中、一印は認識不可の状態を示している。

【0084】

なお、指向性設定装置404においては、発話の開始を示すキーワードとして「カーナビ君」が登録され、発話の終了（図8中の指向性解除コマンド）を示すキーワードとして「ありがとう」が登録されているものとする。

【0085】

初期状態では、指向性設定装置404は全方位の指向性（無指向性）に設定されており、音声認識部403は運転手401及び同乗者402のいずれからも発話開始のキーワードを受け付ける状態となっている。

【0086】

ここで、運転手401が「暑い」と発声するものとする。この発話は、指向性設定装置404内の全方位に設定されているマイクロフォンアレイ（マイクロフ

オン 1 0 1 -1 ~ 1 0 1 -N (図 3 参照) ) を介して抑圧されることなく認識部 3 0 0 (図 3 参照) に供給される。認識部 3 0 0 は発声された単語「暑い」とキーワード「カーナビ君」とを照合する。「暑い」はキーワードと一致しないので、この音声認識結果は認識部 3 0 0 においてリジェクトされる。

## 【 0 0 8 7 】

次に運転手 4 0 1 が「カーナビ君」と発声するものとする。この発声は、全方位の指向性を有するマイクロフォンアレイを介して認識部 3 0 0 に供給される。認識部 3 0 0 は、運転手 4 0 1 が発声した「カーナビ君」が発話の開始を示すキーワードであることを検出すると、発話期間に基づく指向性判定期間を設定して、指向性制御部 2 0 1 に指示する。

## 【 0 0 8 8 】

指向性制御部 2 0 1 は、指向性記憶部 2 0 2 に記憶されている方向データを用いて、運転手 4 0 1 の方向を目的音の到来方向として検出する。指向性形成部 2 0 3 は指向性制御部 2 0 1 からの到来方向設定信号に基づいて運転手 4 0 1 方向に狭指向性を形成する。これにより、以後、運転手 4 0 1 方向からの音響が強調されて、指向性設定装置 4 0 4 から音声認識部 4 0 3 に供給される。

## 【 0 0 8 9 】

ここで、運転手 4 0 1 が「温度を下げて」と発話するものとする。運転手 4 0 1 方向からの音響は強調された高品質の処理音として音声認識部 4 0 3 に供給されており、音声認識部 4 0 3 は、運転手 4 0 1 が発話した「温度を下げて」を正確に音声認識する。この音声認識結果は図示しない車載機器に転送されて、車載機器において各種制御が実行される。例えば、この場合には、車載機器であるエアコンの選定温度が下げられる。

## 【 0 0 9 0 】

次に、車内 4 0 0 の温度が低下したことによって、同乗者 4 0 2 が「寒い」と発声するものとする。しかし、この時点では、マイクロフォンアレイの指向性は運転手 4 0 1 に対して狭指向性に設定されており、同乗者 4 0 2 の発声は十分に抑圧されて音声認識部 4 0 3 に供給される。この場合には、「寒い」の処理音が十分に抑圧されていることから、音声認識部 4 0 3 は「寒い」を音声認識するこ



となく、車載機器において、「寒い」に基づく制御が実行されることはない。

【0091】

また、同様に、同乗者402が温度を上げるためのコマンドである「温度上げて」と発声した場合でも、運転手401に対して狭指向性が設定されており、同乗者402の発声は十分に抑圧されて音声認識部403に供給される。従って、この場合にも、同乗者402の発声に応じて車載機器の動作が制御されることはない。

【0092】

次に、運転手401がマイクロフォンアレイの指向性を解除するコマンド「ありがとう」を発声するものとする。そうすると、指向性設定装置404内の認識部300は指向性形成部203の指向性を全方位に戻す。これにより、認識部300は全方位からのキーワード受け付け状態に復帰する。

【0093】

ここで、同乗者402が「カーナビ君」と発声するものとする。そうすると、マイクロフォンアレイの指向性が今度は同乗者402の方に向き、音声認識部403は同乗者402に対してコマンド受け付け状態となる。更に、同乗者402が、コマンド「温度下げて」を発話すると、同乗者402からの音声は指向性設定装置404によって強調されて音声認識部403に供給され、車載機器のエアコンが制御されて温度を下げることができる。

【0094】

この場合には、運転手401が「暑い」と発話しても、同乗者402方向の狭指向性が設定されているので、音声認識部403には運転手401からのコマンド「暑い」は伝達されない。同様に、認識部300には指向性が形成された処理音が供給されており、運転手401が指向性を解除するコマンド「ありがとう」を発声しても、マイクロフォンアレイによって抑圧されて音声認識部403には伝達されず、発話の権限を同乗者402から奪うことはできない。

【0095】

このように、本実施の形態によれば、発話開始のキーワードを発声した人にマイクロフォンアレイの指向性を向けることにより、周囲の雑音のみならず、同乗

者の話し声も抑圧し、特定発話者の発声のみを正確に伝えることが可能である。また、一度形成された指向性を解除するキーワード（本実施の形態では「ありがとう」）を設定することにより、発話の権限を任意のタイミングで別の人に移すことが可能になる。

#### 【0096】

なお、図1の第1の実施の形態による指向性設定装置を採用した場合には、例えば運転手401に狭指向性が形成されている場合において、指向性を解除するキーワードを用いずに、同乗者402からの発話開始のキーワードにより、新たな同乗者402に対して指向性形成処理を実施することができることは明らかである。

#### 【0097】

更に、図6の第5の実施の形態による指向性設定装置を採用することにより、発話開始のキーワードを発声した発話者方向に対して個別に指向性を形成することができ、2つの指向性形成部によって、運転手401の声と同乗者402の声を分離して音声認識部403に伝達することができる。なお、この場合において、車載装置側で同時処理が困難な場合は、例えば運転手401に狭指向性が形成されている場合において、同乗者402の発話開始のキーワードにより、運転手401に対する処理を止めるか、問い合わせを行い、どちらに指向性を形成するかを選択する等の調停が必要になる。

#### 【0098】

なお、上記各実施の形態においては、指向性設定装置内の認識部と音声認識装置とは別の装置であるものとして説明したが、音声認識装置の認識エンジンを指向性設定装置内の認識部として利用することができることは明らかである。

#### 【0099】

##### 【発明の効果】

以上説明したように本発明によれば、指向性の方向を記憶させることなく、また、指向性の方向を自由に設定可能にすると共に、確実に目的音の到来方向に指向性を設定することができるという効果を有する。

##### 【図面の簡単な説明】

【図 1】

本発明の第 1 の実施の形態に係る指向性設定装置を示すブロック図。

【図 2】

第 1 の実施の形態の動作を説明するためのフローチャート。

【図 3】

本発明の第 2 の実施の形態に係る指向性設定装置を示すブロック図。

【図 4】

本発明の第 3 の実施の形態に係る指向性設定装置を示すブロック図。

【図 5】

本発明の第 4 の実施の形態に係る指向性設定装置を示すブロック図。

【図 6】

本発明の第 5 の実施の形態に係る指向性設定装置を示すブロック図。

【図 7】

本発明の第 6 の実施の形態を説明するための説明図。

【図 8】

第 6 の実施の形態の処理を説明するための説明図。

【符号の説明】

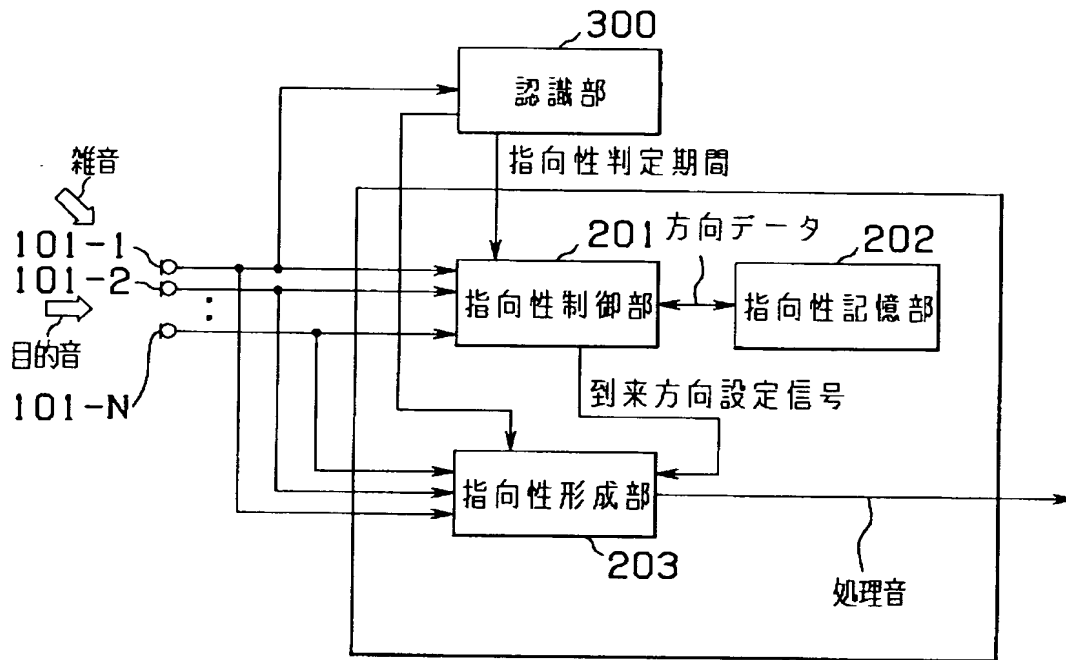
1 0 1 -1～1 0 1 -N…マイクロフォン、2 0 1 …指向性制御部、2 0 2 …指向性記憶部、2 0 3 …指向性形成部、3 0 0 …認識部。

代理人      弁理士      伊 藤      進

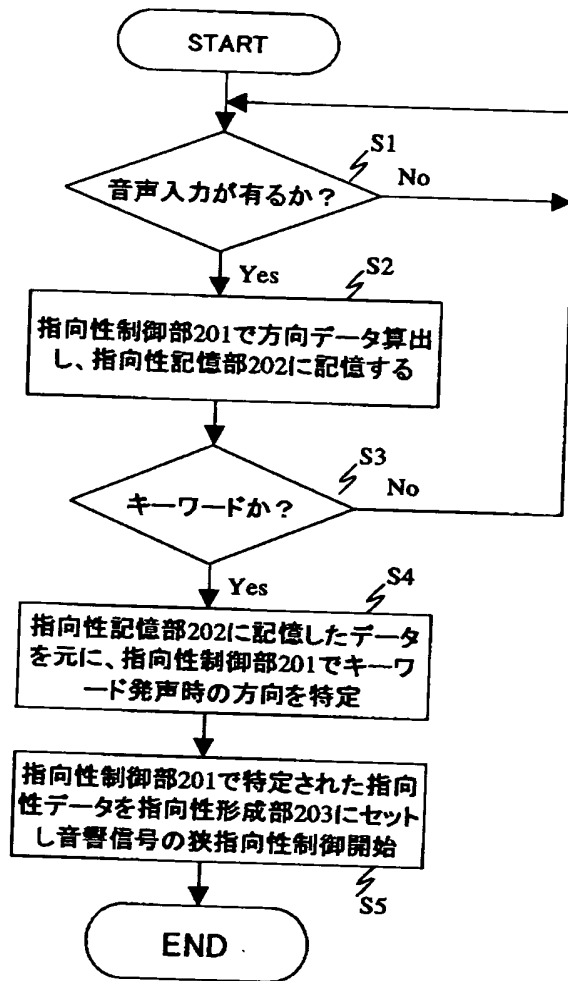
【書類名】

図面

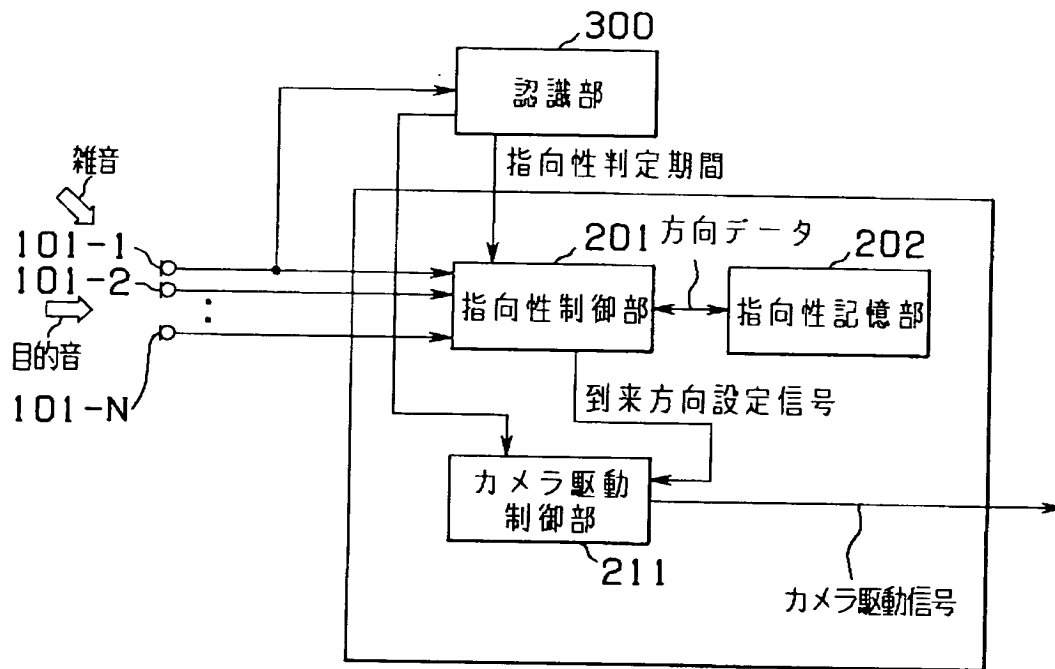
【図 1】



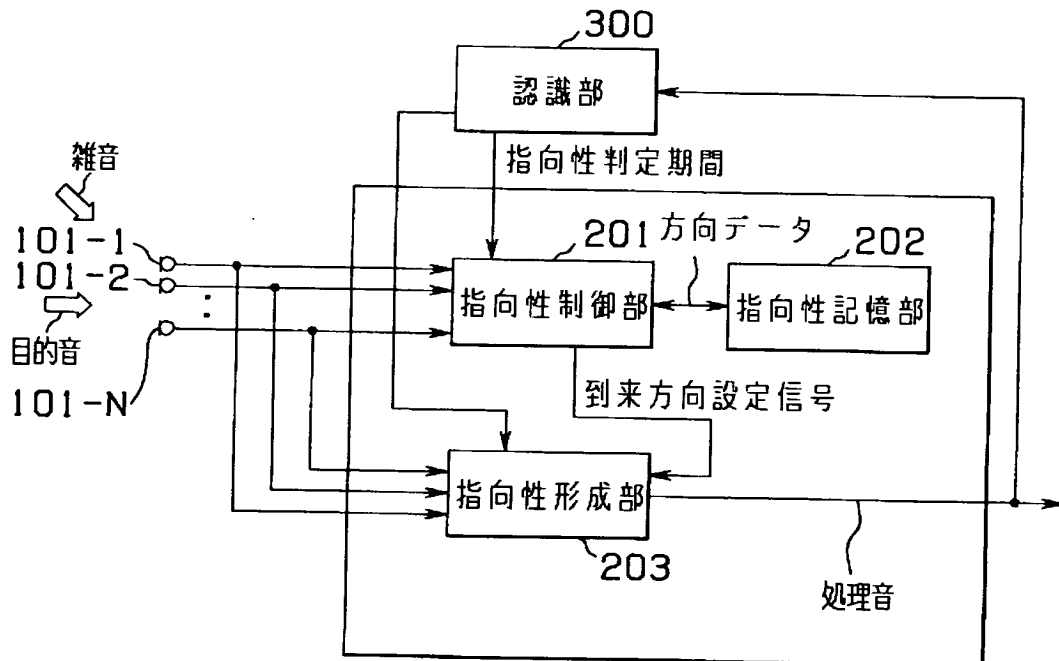
【図 2】



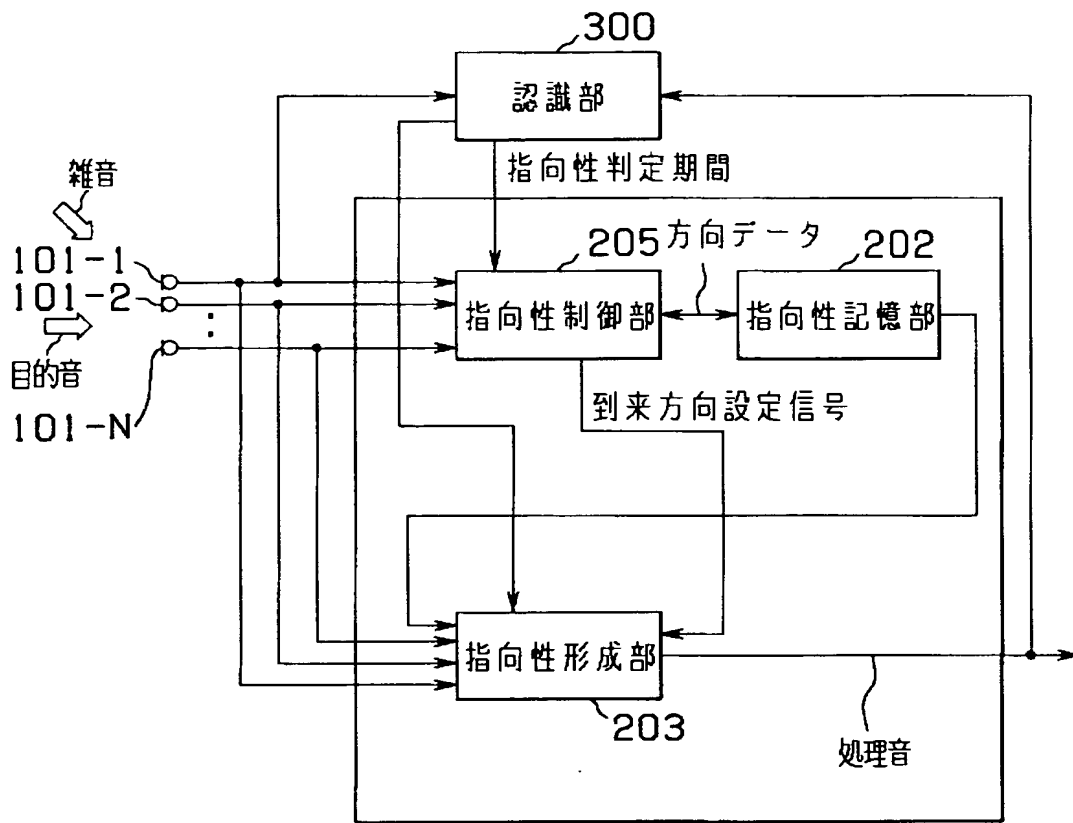
【図3】



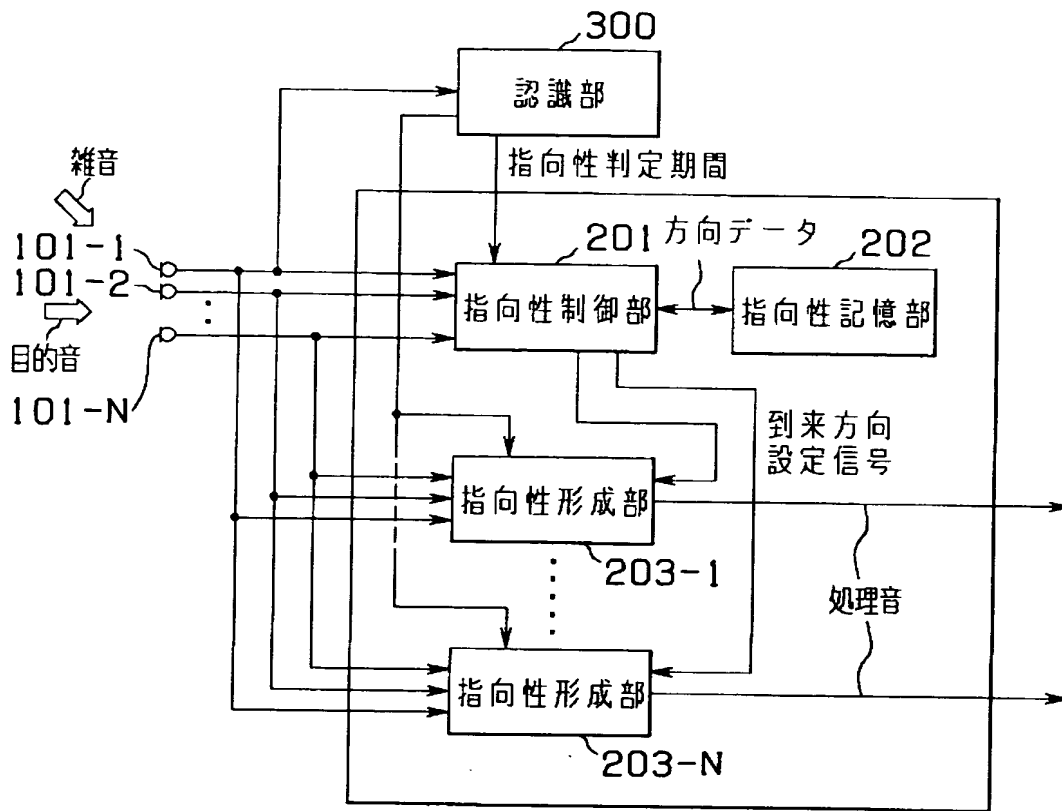
【図4】



【図 5】

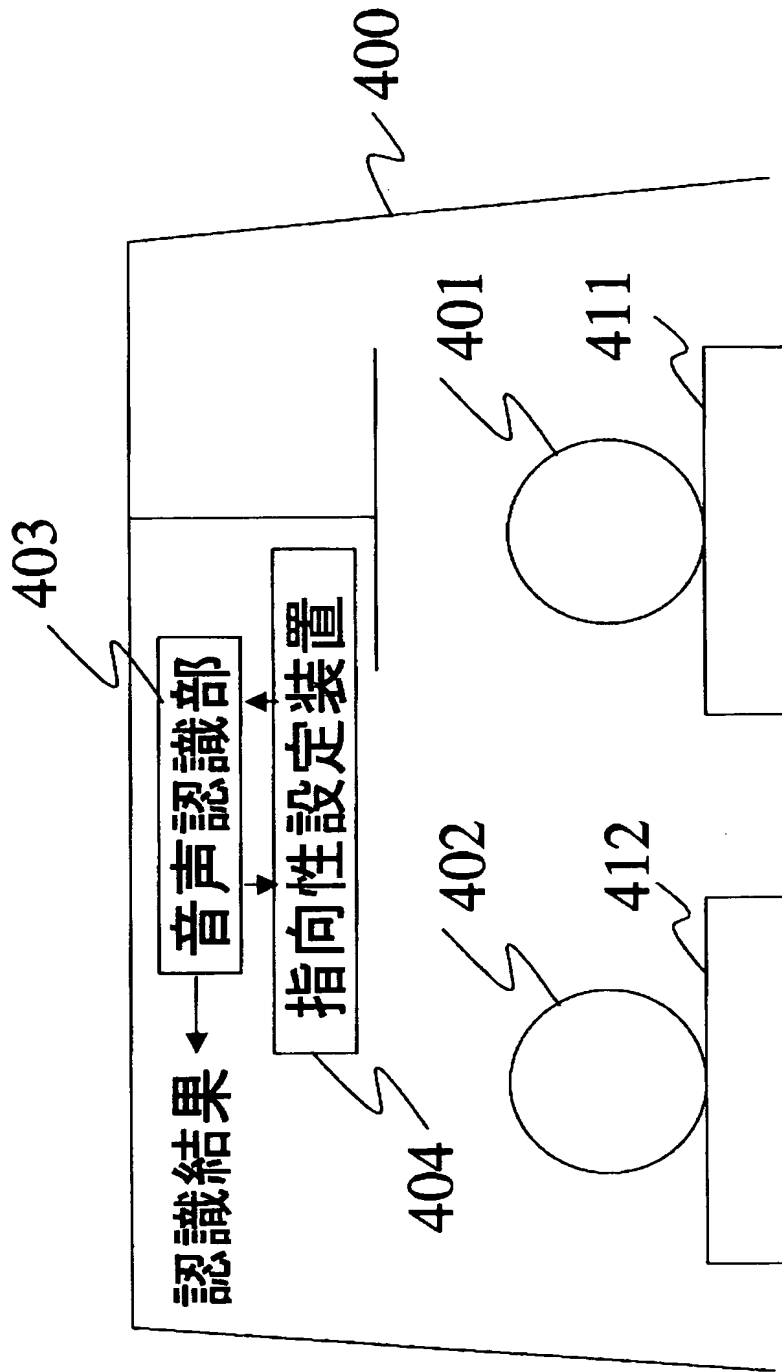


【図 6】





【図 7】



【図 8】

運転手	同乗者	マイクロホンアレーの 指向性の方向	音声認識部 の状態	認識結果
		全方位	キーワード認識	—
暑い		↓	↓	—
カーナビ君		↓	↓	カーナビ君(キーワード)
温度下げて		運転手	コマンド認識	温度下げて(コマンド)
	寒い	↓	↓	—
	温度上げて	↓	↓	—
ありがとう		↓	↓	ありがとう(指向性解除コマンド)
	カーナビ君	全方位	キーワード認識	カーナビ君(キーワード)
	温度上げて	同乗者	コマンド認識	温度上げて(コマンド)
暑い		↓	↓	—
ありがとう		↓	↓	—

【書類名】                    要約書

【要約】

【課題】 固定された方向を登録することなく、自由な目的音の到来方向に確実に指向性を設定することを可能にする。

【解決手段】 認識部 3 0 0 は入力音響からキーワードを検出して、キーワードの発話期間に基づく指向性判定期間を設定する。指向性制御部 2 0 1 は、逐次目的音の到来方向を検出して方向データを指向性記憶部 2 0 2 に記憶させると共に、指向性判定期間における方向データから目的音の到来方向を決定する。指向性形成部 2 0 3 は指向性制御部 2 0 1 によって設定された到来方向設定信号に基づいてマイクロフォンアレイの指向性を形成する。これにより、任意の方向に、1つのキーワードの音声認識によって確実に、目的音到来方向にマイクロフォンアレイの指向性を形成することができる。

【選択図】                    図 1

出 願 人 履 歴 情 報

識別番号 [000003078]

1. 変更年月日	2001年 7月 2日
[変更理由]	住所変更
住 所	東京都港区芝浦一丁目1番1号
氏 名	株式会社東芝